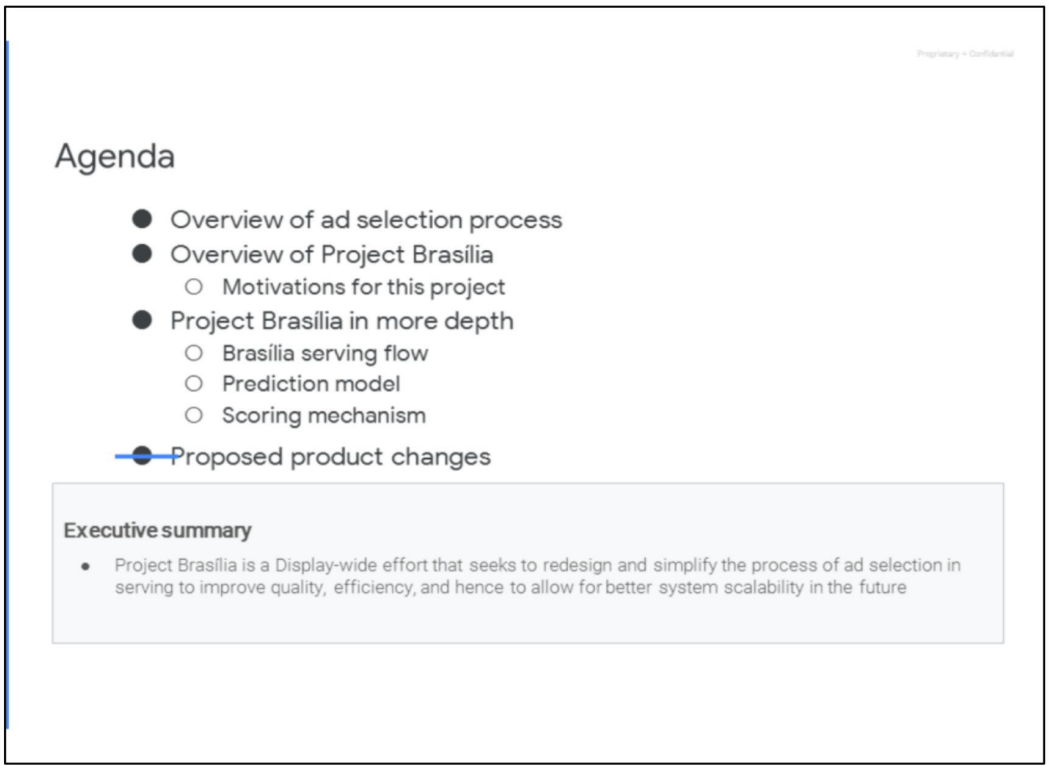

Google Marketing Platform
Brasília: A Unified System for Large-Scale Optimal Ad Selection
DV360 Steering
OCT 13 2020
emaani@



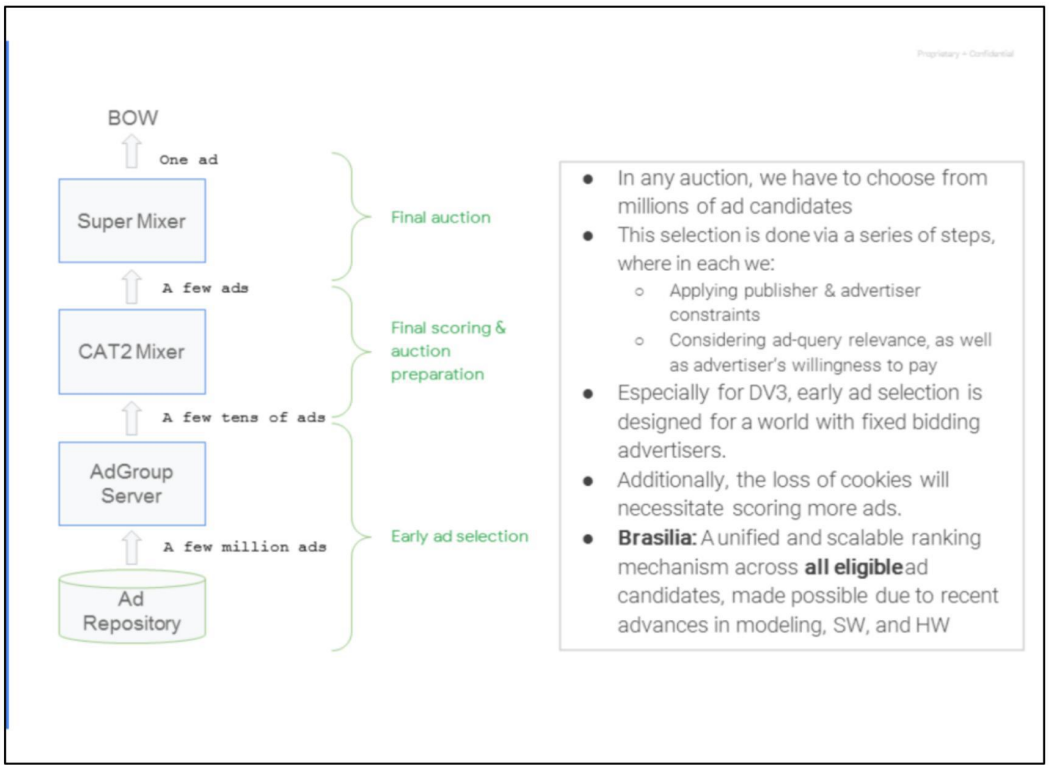
Proprietary + Confidential

Agenda

- Overview of ad selection process
- Overview of Project Brasília
  - Motivations for this project
- Project Brasília in more depth
  - Brasília serving flow
  - Prediction model
  - Scoring mechanism
- Proposed product changes

Executive summary

- Project Brasília is a Display-wide effort that seeks to redesign and simplify the process of ad selection in serving to improve quality, efficiency, and hence to allow for better system scalability in the future


BOW
One ad
Super Mixer
A few ads
CAT2 Mixer
A few tens of ads
AdGroup Server
A few million ads
Ad Repository
Final auction
Final scoring & auction preparation
Early ad selection
In any auction, we have to choose from millions of ad candidates
This selection is done via a series of steps, where in each we:
Applying publisher & advertiser constraints
Considering ad-query relevance, as well as advertiser's willingness to pay
Especially for DV3, early ad selection is designed for a world with fixed bidding advertisers.
Additionally, the loss of cookies will necessitate scoring more ads.
Brasilia: A unified and scalable ranking mechanism across all eligible ad candidates, made possible due to recent advances in modeling, SW, and HW

There are two types of filters that are reducing the ad candidates: 1) enforcing a deterministic business logic (adgroup's targeting, frequency caps, etc) 2) filters that are trying to find the most relevant ads (quality filters). Project Brasilia seeks to unify all those quality filters to use a single principled scoring mechanism.


Ad Repository
Project Brasília Overview
Unified ranking mechanism across all eligible candidates


Motivations for Project Brasília
Advertiser ROI/Revenue gains
Consistency throughout the stack by leveraging investment already made in final scoring (e.g., HDMI)
Score way more ads early on (with a high quality scorer)
Accommodate an LPA world
Targeting restricts expected to be fewer (more expansions/full-auto in a cookieless world)
Current system cannot handle 10X-100X more eligible ad candidates after PSI (targeting)
Core savings
Removing bulky and duplicated scoring and ranking logic
Using the new and highly efficient dot product/embedding-based retrieval technology
Simplicity, easier to maintain and experiment
A more principled logging and experiment framework in AGS
Easier to understand and iterate on by removing heuristics and complex logic
Google
Confidential + Proprietary

More on motivations specifically for DV3

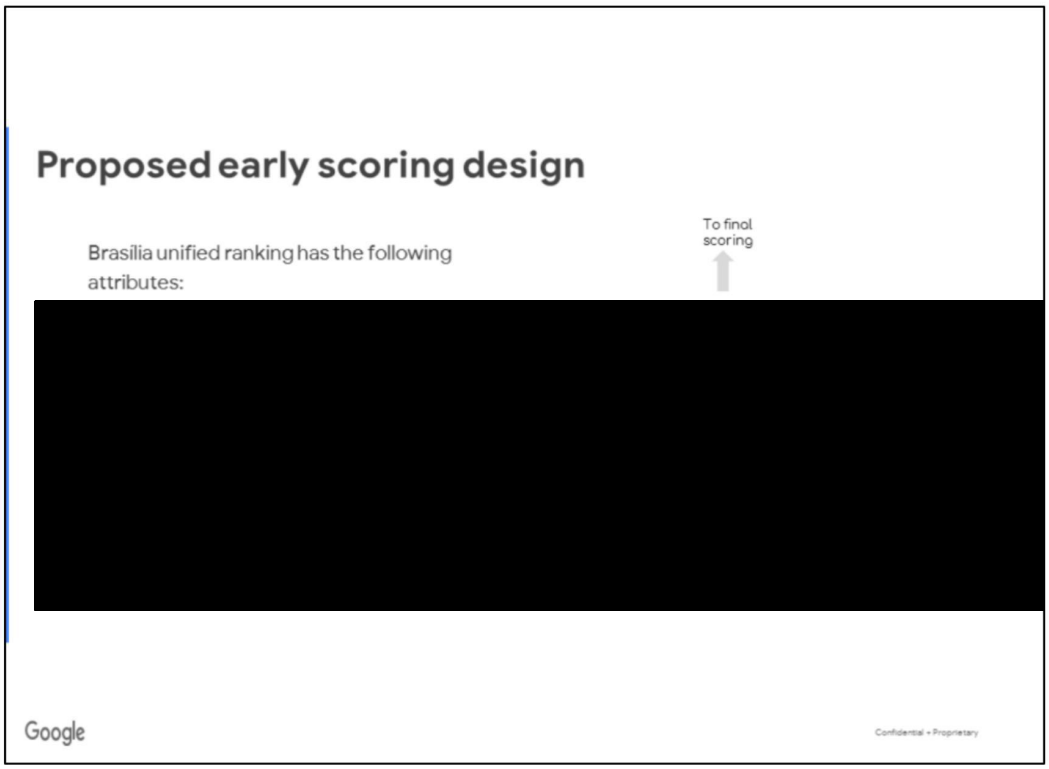
- DV3's current scoring mechanism in AGS is designed for a fixedCPM world
  - An average second-price bid (CPM) is computed per adgroup
  - There are no query-dependent models (unlike in GDA)
  - At the same time since Skyray, auto-bidding adoption grew from ~15% to 80% of eligible spend
- Early ad scoring used to be not as critical due to relatively small ad repository
  - Number of agroups grew from tens of thousands to close to a million since Skyray
    - Monthly active adgroups as of now are ~1M (compared to ~8M in GDA)
- As a result, we hope to see a larger quality gain in DV3 by revamping early ad scoring

Google


Current early ad selection system
To final scoring
Designed to narrow down selection in many stages to reduce computational cost
Some filtering stages are expensive and need to be managed
Includes many heuristics and hand-tuned logic & parameters (e.g., cutoffs)
Scoring happens in at least 4 different stages (highlighted), some are model-based and some are query-independent (e.g., heuristics)
Models have different prediction objectives (eCPM, clicks, conversions, users similarity, etc)
See the full-list of AGS filters here.
Google
Confidential + Proprietary

Id	Date	Text
1	10/13/2020 20:45:40	would similar tunning/heuristics handcrafting be needed to get to a lunch-able version of the project?
1	10/13/2020 20:45:40	we will have num2return cutoff that will be tuned using another model, but that is much less than existing system. On the subject of heuristics, they were added throughout the years, on the new system we try to avoid having any, but we haven't got to the experiment yet to say that is absolutely doable.
1	10/13/2020 22:29:10	Early budget filtering should not be enabled for the XBID corpus, mostly because we haven't figured out how to support Troubleshooting in conjunction with it.

Google

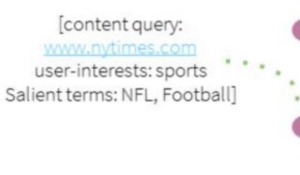

Proposed early scoring design
Brasília unified ranking has the following attributes:
To final scoring
[redacted]
Google
Confidential + Proprietary

Id	Date	Text
1	08/17/2020 05:26:18	While we are at it, any thoughts on exploration? I.e. in the first stage PSI match?
2	08/17/2020 05:26:18	We are thinking of building an "exploration" channel where a percentage of adgroups and creatives are randomly selected and are sent to mixer. This should help with the feedback loop -> if we never select a creative then we don't know it's performance. In this case PSI match only enforces must have business logic that the advertiser asked for, so randomization only happens afterwards in the scorer. Is this inline with what you are thinking or you mean even in PSI match we allow some randomization?

Google

Confidential + Proprietary

Intuition behind factorized models



Ad Candidates	Relevance Score
Nike sports accessories	0.85
Adidas running shoes	0.80
Gieco car insurance	0.60
Rolex luxury watches	0.55
Booking.com hotel ad	0.45

Project ***queries*** and ***ad candidates*** into a single **n-dimensional** space. ***Ads*** and ***queries*** are represented as unit vectors in that space such that "**relevant**" ***ads*** and ***query vectors*** have high cosine similarity. Ads that are ***"irrelevant"*** to a query have vectors that are almost perpendicular to the query vector.

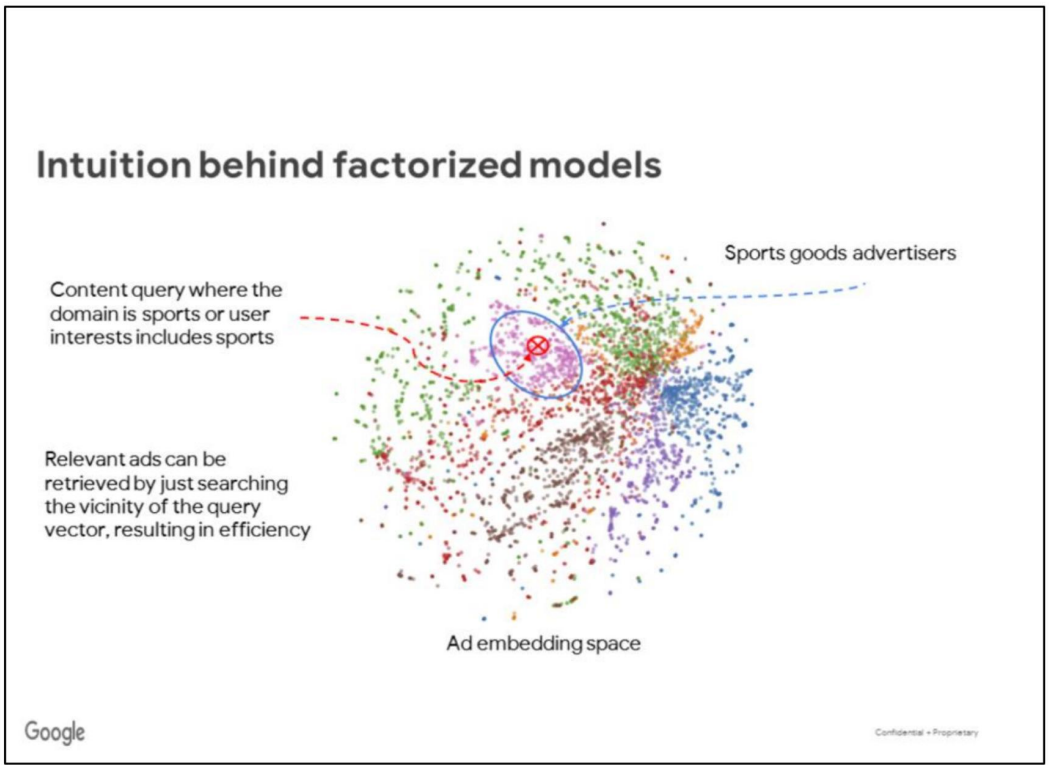
"Relevance" can be defined by probability of user-interaction with the ad (such as a click or a conversion)

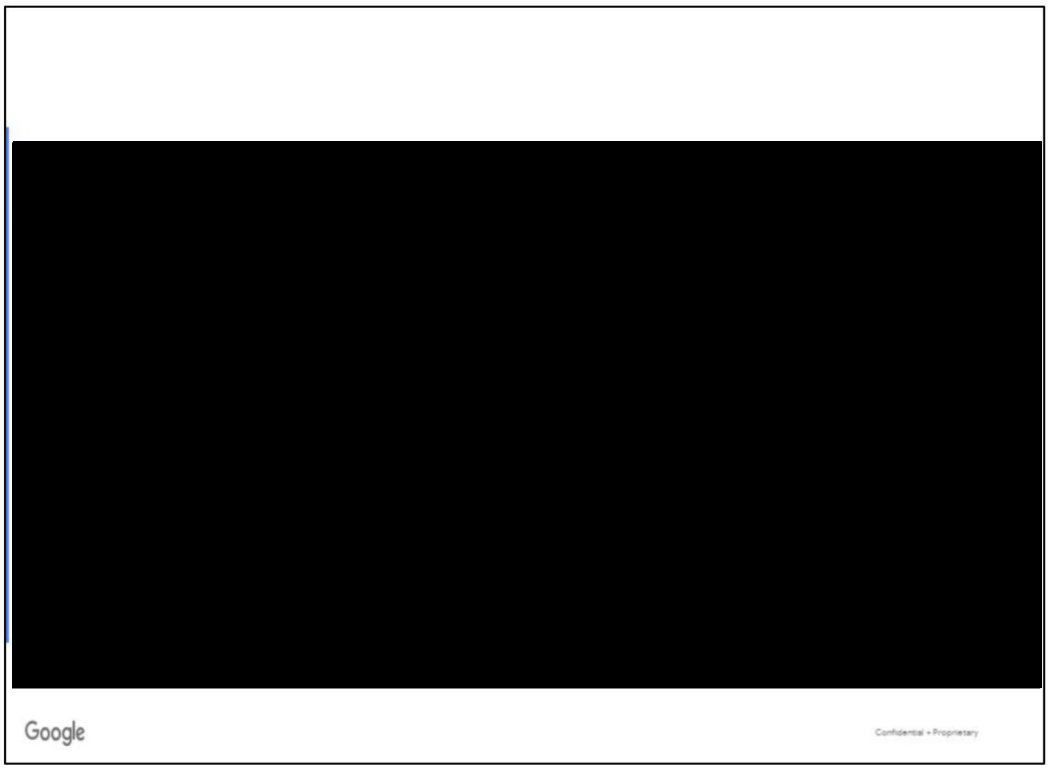
Google

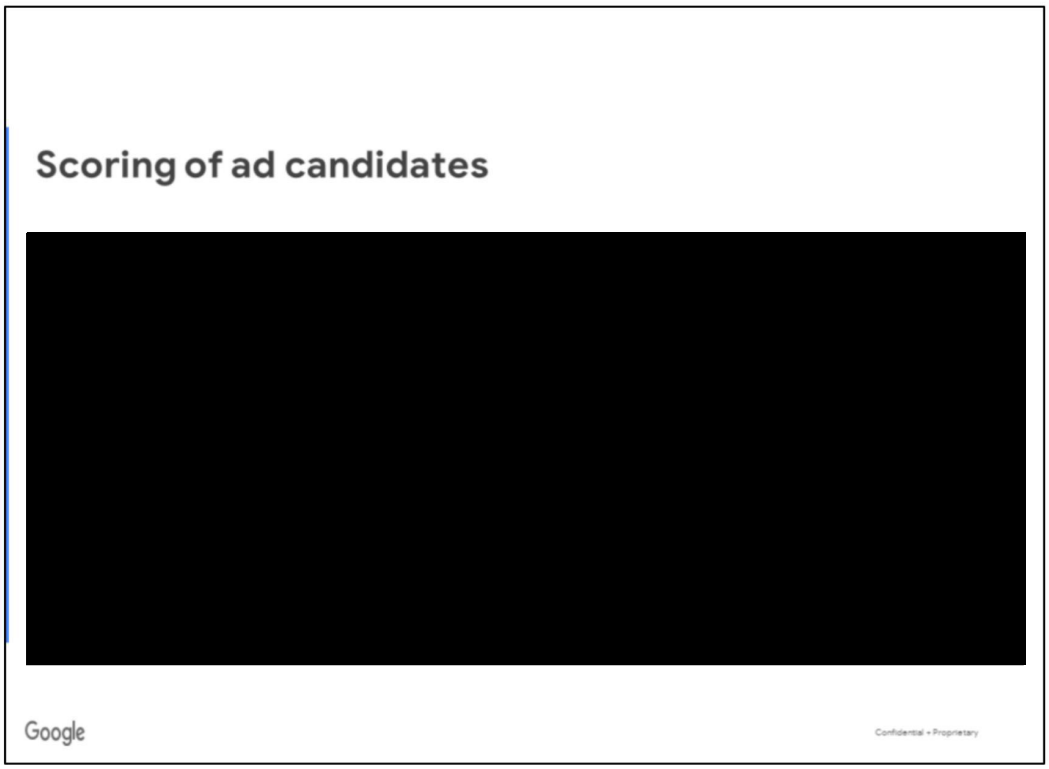
Confidential + Proprietary

Id	Date	Text
2	10/13/2020 19:48:13	do you have a link with a table of how this relevance score would work for different bidding strategies (maximize active views, TOS10/CIVA, custom bidding). How about the AV targeting early filtering?

Google


Intuition behind factorized models
Content query where the domain is sports or user interests includes sports
Sports goods advertisers
Relevant ads can be retrieved by just searching the vicinity of the query vector, resulting in efficiency
Ad embedding space
Google
Confidential + Proprietary


Google
Confidential + Proprietary


Scoring of ad candidates
[redacted]
Google
Confidential + Proprietary

Id	Date	Text
2	08/17/2020 05:28:58	how come there is no component of predicted cvr? What am i missing here?
3	08/17/2020 05:28:58	"Signal" here is essentially a predicted CVR for conversion optimized campaigns. For click optimized campaigns it is predicted CTR, etc. The only difference between this and the pCVR that we have is that the value comes from a dot product of 2 vectors of 128 dimensions each.
3	10/13/2020 19:50:27	what about cold starts? or it will completely rely on HDMI (bidding system)?

Google

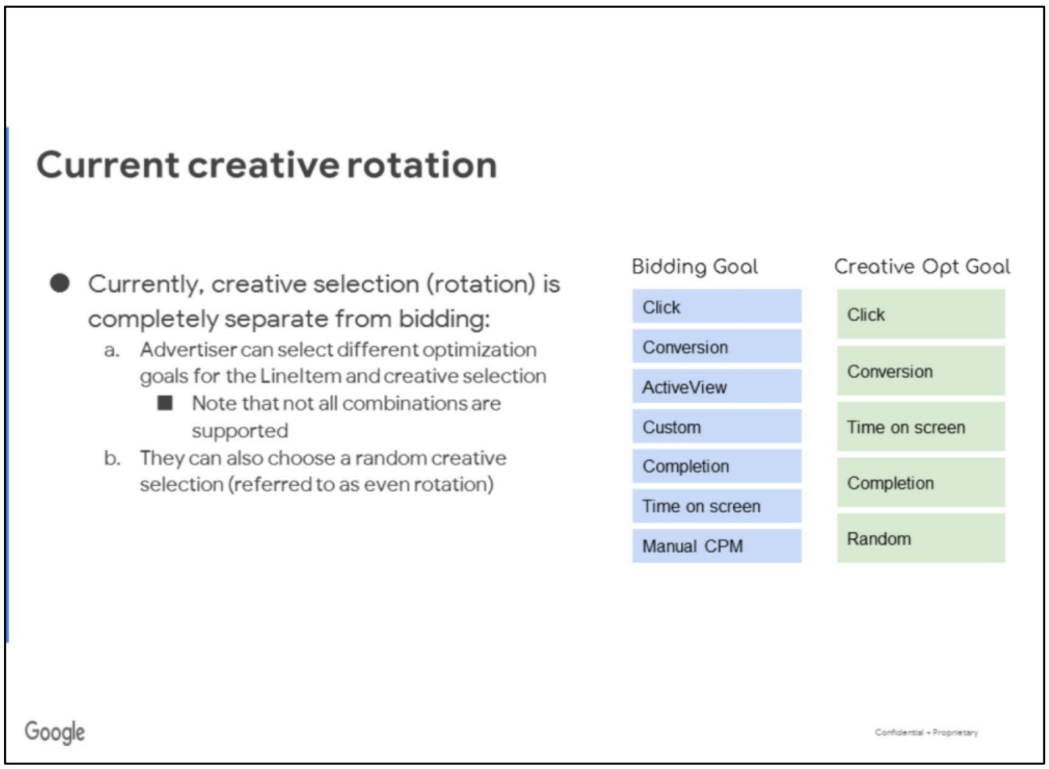
Status

- High-level design is done (exp launch in Q1)
- Created multiple workstreams to work in-parallel on multiple components
  - Modeling: Initial models ready for analysis and experimentation.
  - Serving of models: GDA ready. DV3 needs more work (expected in Q4).
  - Load tests: Initial load test is done for GDA. More on the way especially for bid stack.
  - AGS Infra: Went over all AGS filters to see how they need to change.
  - AGS Auction code: Code complete.
  - Offline HDMI optimization: Developed a new HDMI platform shared between GDA/DV3.
  - Feature research: Work with Doc Targeting and other teams to enable useful and generalizable features for the models.

Google

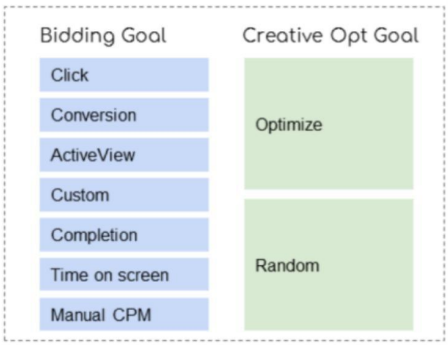

Impact on Creative Rotation
Google
Google

Id	Date	Text
1	10/19/2020 20:19:35	@nekbote@google.com @mkahan@google.com Niranjan - the essence of this proposal as it relates to creative opt is that creative selection can happen at the same time as ad group selection. Moshe and I have met with this team (Nirmal, Ehsan) previously and they are aware of the work underway to have CM be smarter about creative selection by using DV3 rotation strategy to guide it. We've also emphasized need to account for decisioning within CM as well since a lot more creative selection decisioning occurs on CM than DV3 today. They would like to do that in a later phase.


Current creative rotation
Currently, creative selection (rotation) is completely separate from bidding:
a. Advertiser can select different optimization goals for the LineItem and creative selection
Note that not all combinations are supported
b. They can also choose a random creative selection (referred to as even rotation)
Bidding Goal
Click
Conversion
ActiveView
Custom
Completion
Time on screen
Manual CPM
Creative Opt Goal
Click
Conversion
Time on screen
Completion
Random
Google
Confidential + Proprietary

Creative rotation proposal

- Brasília can do adgroup and creative scoring in a single step. To be able to achieve this:
  a. Advertisers can choose between two options (to align with GDA):
     - Optimize creative selection towards the goal of the LineItem
     - Randomly select creatives
  b. For fixed bidding LineItems we would need to ask for a "goal" (e.g., clicks, views, conversions)
     - This goal can be used for creative selection and perhaps other "optimization" features



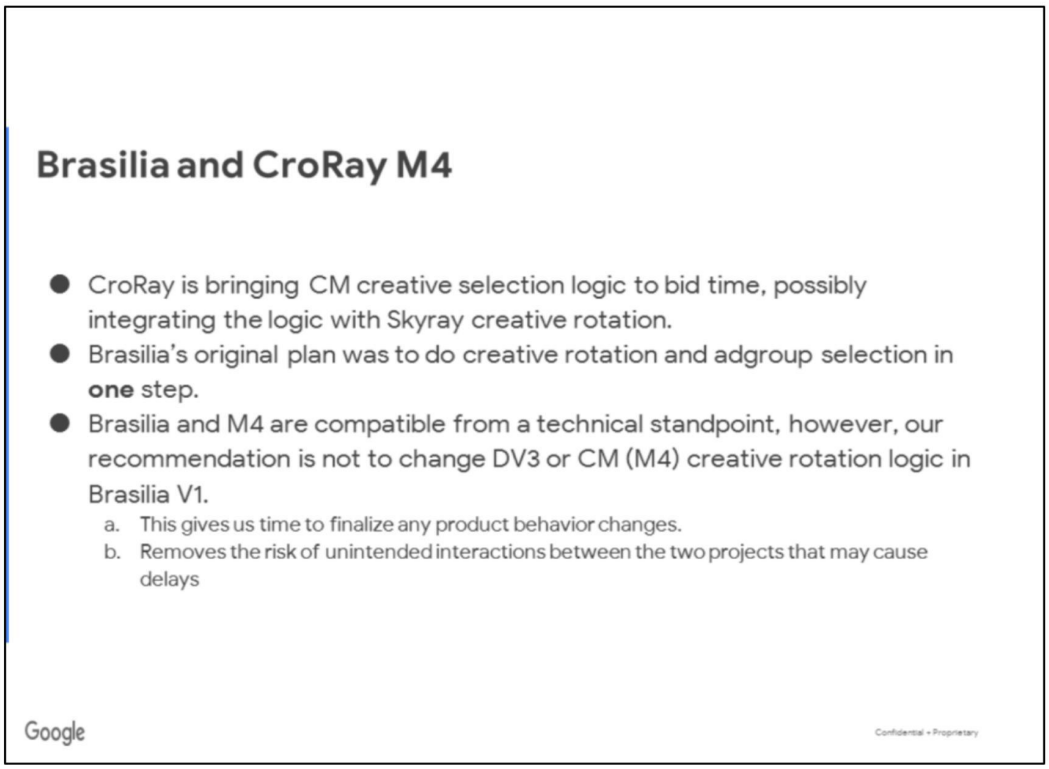
Google

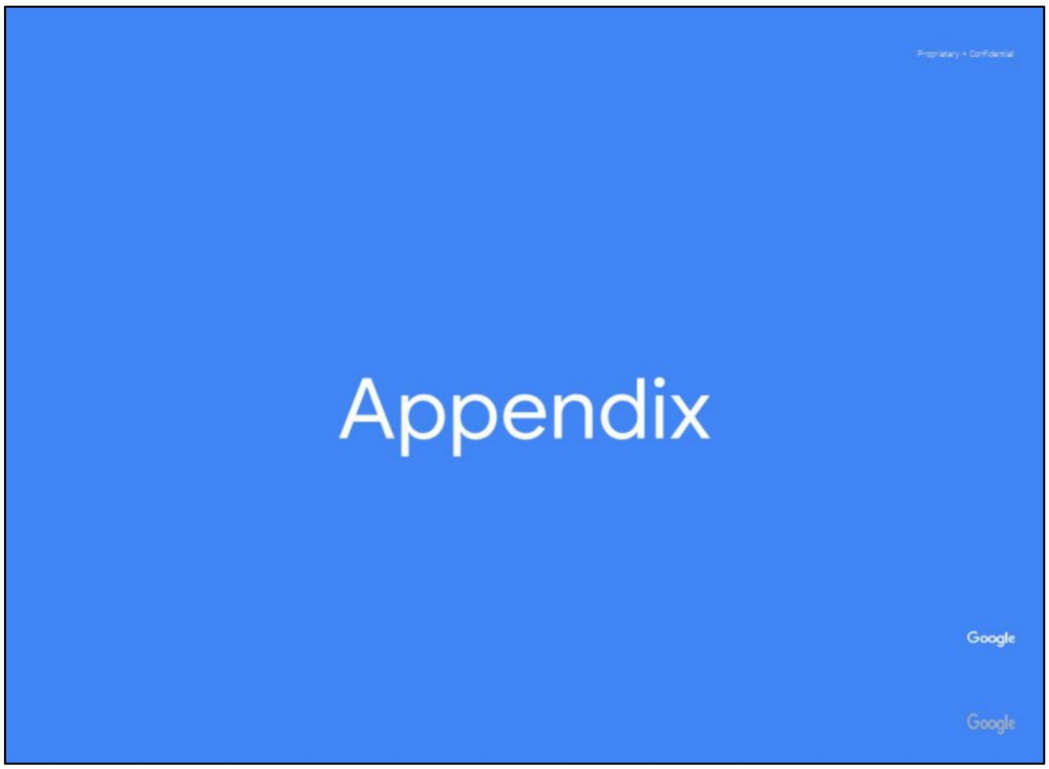
Confidential + Proprietary

Id	Date	Text
2	07/14/2020 13:40:19	There is a concept of IO goal and that is 'mandatory' to express for every IO. The quality of that goal is not always accurate, but its the best we have today for fixed bidding. https://screenshot.googleplex.com/qUG8Zh3FHaK.png
1	07/14/2020 13:41:58	Do we want to call out anything with regard to CM? If most creative optimization decisions are made in CM today, we may want to address this since if we don't the impact will be limited.

Google

Confidential + Proprietary


Brasilia and CroRay M4
CroRay is bringing CM creative selection logic to bid time, possibly integrating the logic with Skyray creative rotation.
Brasilia's original plan was to do creative rotation and adgroup selection in one step.
Brasilia and M4 are compatible from a technical standpoint, however, our recommendation is not to change DV3 or CM (M4) creative rotation logic in Brasilia V1.
a. This gives us time to finalize any product behavior changes.
b. Removes the risk of unintended interactions between the two projects that may cause delays
Google
Confidential + Proprietary


Proprietary + Confidential
Appendix
Google
Google

Silo vs Shared Models

- In mixer, we continue to build per-advertiser silo models to compute first-price optimized bids
- In AGS, we are exploring two options to compute advertiser's Desired Spend:
  - A single factorized model that predicts **views**, **clicks**, and **conversions**
  - Three separate models that predict **views**, **clicks**, and **conversions**

- **Question:** for the early stage, we have legal approval to build a shared model, but we are looking to discuss more broadly.

Google

Desired Spend Example

Advertiser 1: Wants to buy conversions at $12CPA. targeting Bay Area

Advertiser 2: Wants to buy conversions at $10CPA. targeting California

Today (in any internal auction):

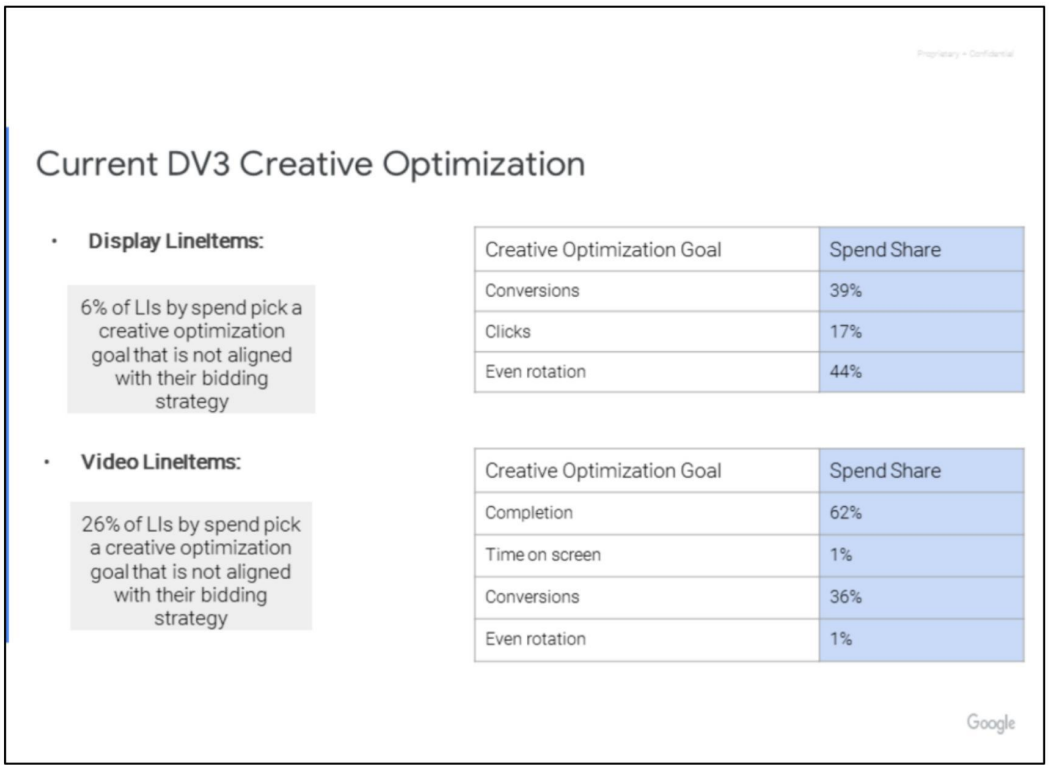
Say Auction discount is 50% in California, 20% in Bay Area. and assume CVR is 10% for both.

We calculate maxCPM1 = 12 * 10% / (80%) = **$1.5**, maxCPM2 = 10 * 10% / (50%) = **$2.0**.

We allocate this auction to Advertiser 2.

But allocation to Advertiser 1 results in more revenue.

Google



Current DV3 Creative Optimization

- **Display LineItems:**

6% of LIs by spend pick a creative optimization goal that is not aligned with their bidding strategy

Creative Optimization Goal	Spend Share
Conversions	39%
Clicks	17%
Even rotation	44%

- **Video LineItems:**

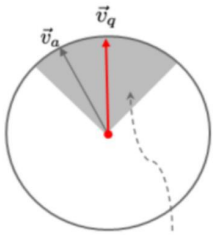
26% of LIs by spend pick a creative optimization goal that is not aligned with their bidding strategy

Creative Optimization Goal	Spend Share
Completion	62%
Time on screen	1%
Conversions	36%
Even rotation	1%

Google

Modeling Intuition

In 2-Dimensional space, query and ad vectors are projected into a unit circle.



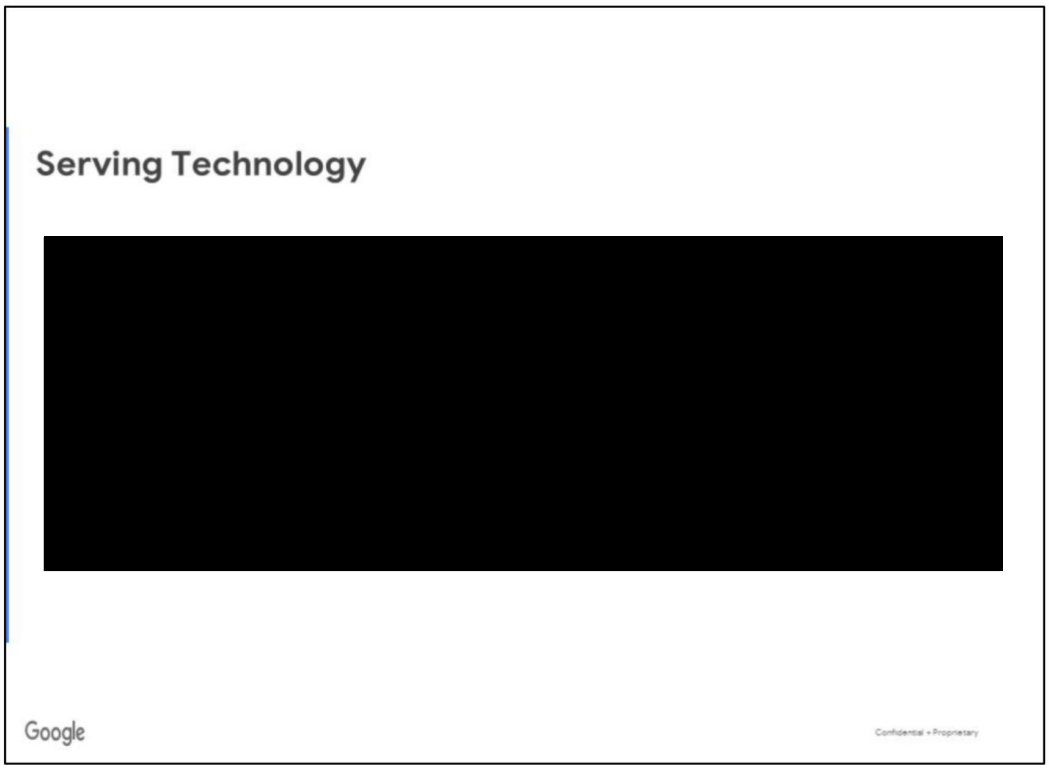
Probability of Cosine similarity score > 0.7 in 2-D is **0.25**

In 128-Dimensional space, query and ad vectors are projected into a unit sphere.

Probability of Cosine similarity score > 0.7 in 128-D is **2.7e-20**

Google

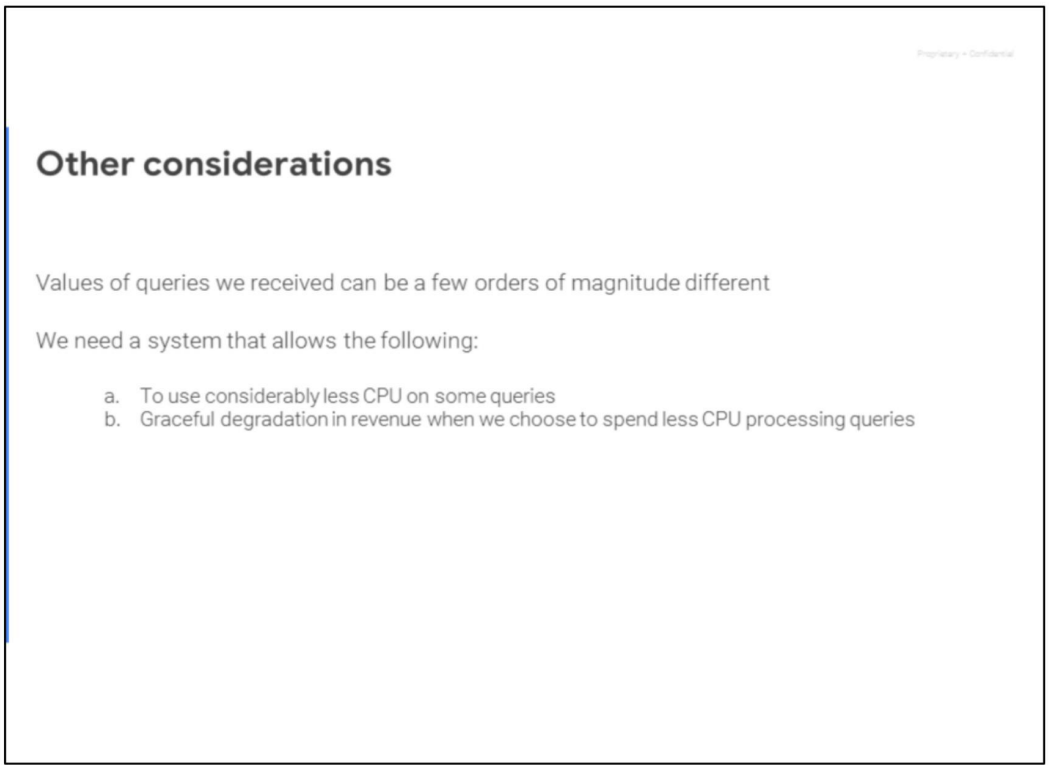
Confidential + Proprietary


Serving Technology
[redacted]
Google
Confidential + Proprietary

Id	Date	Text
4	10/13/2020 20:33:02	what would happen if this fails?
4	10/13/2020 20:33:02	we have query-level and query independent fallbacks if this fails.

Google

Confidential + Proprietary


Other considerations
Values of queries we received can be a few orders of magnitude different
We need a system that allows the following:
a. To use considerably less CPU on some queries
b. Graceful degradation in revenue when we choose to spend less CPU processing queries

Brasília & LPA

- In today's world, **RMKT** and **AUDIENCE** targeting significantly reduce number of ad candidates that need to be scored by helping advertisers narrow down their audience
- In a world, where our ability to effectively use those tools to narrow down advertisers' audience is limited, we need a more effective, consistent, and scalable way to score and select ads using the optimization technology
- We plan to run an experiment (1%) to test **Brasília's** effectiveness in this world:
    - For non-remarketing ads, in the experiment arm, expand targeting to Run-of-Network (RON) and use Brasília scoring to select most relevant ads